
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

PEAT- a tool to predict team performance in systems

PLEASE CITE THE PUBLISHED VERSION

<http://dx.doi.org/10.1109/SYSOSE.2010.5544047>

PUBLISHER

© IEEE

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Sinclair, Murray A., Carys E. Siemieniuch, Roger Haslam, Michael Henshaw, and L. Evans. 2019. "PEAT- a Tool to Predict Team Performance in Systems". figshare. <https://hdl.handle.net/2134/7230>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

PEAT- a tool to predict team performance in systems

M.A. Sinclair

Centre for Innovative & Collaborative Engineering
Loughborough University
United Kingdom
m.a.sinclair@lboro.ac.uk

R.A. Haslam

Dept. of Ergonomics
Loughborough University
United Kingdom
r.a.haslam@lboro.ac.uk

C.E. Siemieniuch

Dept Electrical & Electronic Engineering
Loughborough University
United Kingdom
c.e.siemieniuch@lboro.ac.uk

M.J.d.C. Henshaw

Dept Electrical & Electronic Engineering
Loughborough University
United Kingdom
m.a.sinclair@lboro.ac.uk

L. Evans

Human Factors Dept
BAE Systems Advanced Technology Centre
United Kingdom
laird.evans@baesystems.com

Abstract - *The paper describes the development of a tool to predict the success of a team in executing a process. It is expected to be used by systems engineers in initial stages of systems design, when concepts are still fluid, including the structure of the team(s) who are expected to be operators within the system. Currently, the tool is undergoing verification and validation; to date, the tool predicts fairly well and shows promise.*

Keywords: Systems Engineering, Performance Prediction, Organizational Performance

1 Introduction

The original purpose of the tool was to provide designers of military systems with a simple tool to use in the conceptual stages of design (when variables are still variables and not parameters) to help in risk reduction exercises when considering the staffing of processes. Some sample questions for which the tool could help in providing answers are:

- What is the likelihood that this team will be successful in executing the given process? *Given appropriate inputs (see below), the tool produces a likelihood of success.*
- By how much can the team size be reduced, before the likelihood of success becomes unacceptable? *Stepwise, the team size would be reduced, and for each step the duties of the team would be re-arranged and the tool would be used to recalculate the likelihood of success.*
- By how much can the attributes of the individuals in the team be reduced, before the likelihood of success becomes unacceptable? *Progressively, the ratings for each team member would be reduced, and for each reduction the tool would be used to recalculate the likelihood of success.*

- Which of these two (three, ...) proposed teams is likely to be the most successful in executing the process? *Each team arrangement would be assessed for success by the tool; either the user could select the 'best' team, or the user could examine the trends in the likelihoods of success, and devise an even better team*
- Can an acceptable likelihood of success be achieved by increasing the quality of the individuals in the team? *Progressively, the ratings for each team member would be increased, and for each increase the tool would be used to recalculate the likelihood of success.*
- Can an increase in automation of tasks, or a re-arrangement of tasks, in the process bring the team's likelihood of success up to an acceptable level? *For each step in the re-ordering or re-classification of tasks, the tool could be used to calculate a likelihood of success.*

It has subsequently become evident that these questions could also be developed within the context of Through-Life Capability Management (for systems), with regard to the delivery of Service Level Agreements (guaranteeing performance of the system).

It will also be noted that these questions are phrased in terms of success. This is a significant point; 'success' is defined here as executing the process correctly and attaining all of the goals of the process; no reworking, no extra resources, no extra time. 'Likelihood', as usual, is expressed as a probability. Note that $(1 - p(\text{success}))$ does not mean 'probability of failure'; for many processes, if they show signs of failure, more resources, or more time can be allocated, and the process can be brought to some degree of successful completion.

Note also, that project success does not mean victory; the team might fire the surface-to-air missile successfully,

but it might still miss the target. 'Success' solely refers to the process execution.)

2 Development of the tool

From the outset, three constraints determined the development of the tool:

- A user-defined constraint: If the tool needs more than two pages of A4 to explain it, it will not be used. This was a unanimous view among a target group of engineers who were interviewed at the beginning of the project.
- A business process constraint: Systems designers in the UK are already familiar with, and may be using, techniques such as HEART and CREAM for assessing the reliability of individuals. The tool should incorporate these techniques, or any other in-house technique, to enhance ease of acceptance into design processes.
- A design constraint: At the conceptual stages of design, little will be known about the individuals in the team that executes the process, apart from generic attributes. Equally, the process will be undefined – perhaps just a single flow diagram sketched on a sheet of A4. Hence, the tool must make a minimum demand for input data.

The first constraint has been met, without recourse to minuscule script; a second, longer manual has also been provided to explain not just what to do, but why it needs to be done. The second constraint has been met by creating a three-stage tool; Stage 1 collects data about the attributes of the individuals in the team and the intercommunications necessary for execution of the process. Stage 2 collects an analysis of the process environment, using either HEART [1] or CREAM [2], or the organisation's in-house technique. Stage 3 convolves the outputs of Stages 1 and 2, and produces a likelihood of success, depending on the binding of the team to the process.

The third constraint has been met as follows. Input data on individuals cover five variables. The first is an identifier for each person, to aid the user to understand the intermediate steps. The second is a trustworthiness variable; the dependability of the person to deliver results. The third measures team skills; how constructive the person is in aiding the team to its goals. The fourth assesses the knowledge and skills that the person brings to the process being considered. The fifth variable assesses the authority of the person within the team; this covers the third constraint.

In addition, the proposed communication structure (who talks to whom) is captured, as a matrix of one-way links.

It should be noted that neither the team nor the process need to be known in detail. Minimally, a task sequence diagram on a single sheet of A4 paper is

sufficient. Similarly, the roles can be classes, such as Lieutenant, Sergeant, etc. Who does which task needs to be known, as does the communication pattern for the process, but these could be early estimates, with the expectation that the analysis will be revisited later when more information is available.

From the trustworthiness, team skills, and knowledge ratings, the tool produces a Performance Shaping Factor (PSF). High-performance people will have a PSF less than 1.0; for poor performers it will be greater than 1.0.

Next, the user analyses the process environment for each individual. This allows for distributed teams; if the team is co-located, only one analysis needs to be performed. This analysis may be performed using either HEART [1] or CREAM [2], or with an in-house technique substituted for these. Both techniques are used for individuals; HEART asks the user to select a base $p(\text{error})$ for a given task, and then allows the user to modify this by means of Performance Shaping Factors. CREAM asks the user to complete 9 scales, and on the basis of these calculates a $p(\text{error})$. The probabilities of failure that emerge from this analysis are now multiplied by the PSFs from Stage 1 to produce a probability of error for each individual in the team in his/her environment.

Stage 3 happens without necessitating user input. It now combines the values obtained above, the authority ratings, and the communications matrix to arrive firstly at what is called 'interactive probabilities of error' for each individual; in other words, acknowledging peer effect. The initial probabilities of error from the paragraph above are now adjusted by these extra variables. The assumption here is that a person's performance will be influenced by the performances, knowledge, and teamworking capabilities of those who communicate with that person in the execution of the process. This might be feedforward, feedback, or the observable quality of work.

An algebraic equation performs this step, with the property that if a person communicates with very good people, then that person's probability of error will decrease, and vice versa. It does this by first calculating the Power Distance between a person and a peer, then multiplying this by the peer's $p(\text{error})$, then combining the resultant for all communicating peers, and then averaging this latter resultant and the person's $p(\text{error})$.

Hence, we now have an 'interactive probability of error' for each team member.

Secondly, the team's binding to the process is addressed. It was discovered that three classes of binding cover a wide range of real teams. The three classes are described below:

- The 'aircrew' team. Consider a helicopter, flying from A to B over hostile territory. The pilot executes the

process of flying; the rest of the crew act only as advisors (e.g. 'A SAM has been fired'), but do not play a part in flying the helicopter.

- The 'Boatcrew' team. Consider a rowing eight (nine, with the coxswain). From start to finish, each person has a specific task, and cannot perform anyone else's task. Consequently, the absence of any crew member ensures failure.
- The 'Omnicompetent' team. Here, anyone can perform anyone else's task, and may do so in executing the process.

Likelihoods are presented for each of these bindings together with estimated 95% confidence limits for the likelihood; the user may choose the best-fitting example, or may calculate a specific value from the 'interactive probabilities of error' using the laws of probability, if the three classes are deemed unsuitable.

3 Current state of the tool

Due to the adoption of a form of Rapid Prototyping as the development methodology, the tool is now in EXCEL spreadsheet form, in its ninth revision. Functionally, the tool is considered complete save for further verification and validation, and is being made more presentable and understandable to the intended user. The user manuals have also been tested, and are considered complete. In accordance with the first constraint above, a two-page version and a more detailed version have been developed.

At the moment, the tool is constrained to deal with teams of ten or less members. This is a developmental constraint; it is expected that when the tool is deemed to have been fully tested, this constraint will be removed. With cognizance of this constraint, the tool appears to be able to deal with a number of different scenarios; as shown later in the verifications, the teams considered range from military teams to system development teams, to the Board of Trustees for a pension fund; in other words, teams of many different organisational cultures. It can accept distributed teams, and teams where some members do mainly physical tasks and others do mainly cognitive tasks.

4 Verification and validation of the tool

In other words, 'Does it work?'. The evidence so far is that it does. A second question is, 'Is it sensitive enough?'. Again, the indications are that it is. This sensitivity issue is discussed first below, followed by the verification and validation results.

4.1 Sensitivity of the tool.

For a team of two, using CREAM to categorise the environment, 1360 values may be obtained between 0.0 and

1.0. For a team of 10, using HEART the range of values is well over 1 billion. There appears to be sufficient sensitivity for most purposes.

4.2 Verification of the tool.

To date, over 400 verification tests have been carried out on these teams, changing at least one variable for each test. Three classes of teams have been investigated in some detail. Two are teams of four; the difference between them is in the communication patterns. In a 'Linear' team, the members are arranged linearly, with two-way communication between adjacent members (similar to a production line). In a 'Cocktail' team the arrangement is that of a star, with a central person communicating with all the others. The others have partial communications between them; the net effect is that each team member has a different number of communication links. The third class has explored different sizes of teams and different communication patterns, including the effect of two co-operating teams.

A further restriction on all these tests is that all the tests have been performed using CREAM to characterise the environment, since it has better theoretical underpinnings than HEART.

From all of these tests, it seems that some general conclusions can be drawn :

1. Given that the tool has been constructed as a simple technique, with no feedback loops and no 'if-then' rules, it is gratifying that the behaviour is as expected; trends are consistent, and no test produced a prediction outside the range 0.0 – 1.0.
2. It is striking that in all sets of tests, a well-designed working environment (i.e. well-designed process, good HMI, proper training, good support, leadership, etc.) is the biggest contributor to p(success). The other variables in the tool become more important as the working environment degenerates, but they cannot make up for it. In a good working environment, team variables are relatively unimportant; even when the team is of poor quality, $p(\text{success}) > 0.9$. For a military environment, this is equivalent to saying that if all the US DoD Systems Engineering requirements have been met completely and in full (or in the UK, the Defence Lines of Development have delivered in full), taking full account of human factors integration, then successful performance of the process is assured (note that this is not saying, 'Victory is assured' – see earlier comment).
3. In poor working environments, the most important variable is the quality of the team members. As long as there are several high quality members in central roles, able to communicate with the other members, then a level of performance ($p(\text{success}) \approx 0.25$) can be achieved. However, even with a high quality team, it is not possible to lift $p(\text{success})$ above 0.35.

4. A good team is always better than 4 individuals.

Of course, the truth content of these statements depends on the validation studies below, and more of these must be carried out. But, insofar as there is truth in these statements, findings 2 and 3 together have a significant corollary; they provide a strong argument for the importance of Human Factors/ Ergonomics in systems engineering projects. One might imagine that, given the extent and complexity of military systems being sourced for the NATO forces for future decades (particularly the USA), and the likely manning issues for these, these conclusions will be of some interest to those in military procurement, and to those in civilian safety-critical systems management.

4.3 Validation of the tool.

To date, 18 validation exercises have been carried out, summarised in Table 1. All are historical cases, with the results known; i.e. validation-by-criterion. However, in all cases but the last the criterion was subjective, since the processes were not repeated. The last was a repetitive process, allowing comparison of probabilities of success. Extremely brief descriptions of the cases are included for reasons of space and confidentiality, together with the predicted success and the source's verbatim comment on the result. Note that in all cases except for two, the respondent in each of the tests was a Systems Engineer.

Table 1: Aggregated data for validation studies

#	Team process	Prob success	Comment on result
1	Insert new FCS in tank	0.76	"That looks good."
2	Dev't of control system for UAV	0.59	"OK, if a little generous"
3	Dev't of systems Health Mgmt System	0.87	"I'm happy with that result - perhaps a little bit high"
4	Create engineering Technical Demonstrator	0.92	"Result is OK; perhaps a bit high"

#	Team process	Prob success	Comment on result
5	Dev't of comms system for navy ship	0.95	"Rings reasonably true"
6	Bid prep'n for US DoD ITT	0.45	"Result looks OK; wouldn't want to argue with it; perhaps a little bit high."
7	Execution of a Des & Build project for M.Eng degree	0.49	"Result is a bit low"
8	Deliver HFI to manufacturing Technical Demonstrator	0.92	"OK; but doesn't account for a weak team member. Got it together because of the efforts of the rest of the team"
9	Dev't of guides for UK Gov't dept.	0.90	"That's OK"
10	Software development for UK NHS	0.60	"Estimate is a bit low; would have expected about 0.75"
11	Development of a UAV ground station	0.48	"A bit low - would have expected around 0.7"
12	Management team in University	0.83	"Result is OK, but this isn't a normal team"
13	Pension Bd of Trustees managing fund	0.99	"That's about right"

#	Team process	Prob success	Comment on result
14	SAS patrols in hostile territory	0.63	“Difficult to assess ; if contact with the enemy, all plans change, therefore ‘failure’. But no contact = failure. Overall, value looks good.”
15	Mentoring team for military Outward Bound scheme	0.90	“That’s OK; we got it right most of the time despite outside influences”
16	Preparing armoured vehicles for Gulf War	0.89	“On average, 14 out of 16 would go straight through.”
17	Preparing RAF Tornados for Gulf War	0.082	“About right – only 3 of 26 went through without rework”
18	High-tech jobbing shop making military-standard RAM for development studies	0.97	“That’s interesting. Expected monthly performance for this process is between 0.90 and 0.98.”

A Kolmogorov-Smirnoff test [3] was carried out to discover any significant departure from accuracy (note that a Chi-squared test is not appropriate due to Cochran’s criterion [4]). This is shown in table 2 below. Getting 3 predictions wrong represents a proportion of 0.167. According to table E of Siegel [3] this indicates that the Null hypothesis (no departures from Ideal) is not rejected ($p > 0.05$).

Table 2: Data for Kolmogorov-Smirnoff test, from Table 1

Classes	Ideal cumulative	Actual cumulative	Prop. difference
Wrong prediction	0	3	0.167
Correct prediction	18	18	0

A further, Binomial test [3] was executed on those estimates deemed above and below, to test for a bias in the predictions. Table 3 shows the results of this.

Table 3: Data for Binomial test, from Table 1.

Prediction below user’s opinion	Prediction above user’s opinion	Predictions deemed correct
3	4	11

According to Table D of Siegel [3], this indicates that the Null hypothesis (no bias in predictions) is not rejected ($p > 0.05$).

These tests indicate that the tool can produce reasonable predictions, and therefore, the inferences drawn from the verification tests are trustworthy. However, it has been arranged that three further independent, detailed validation tests will be carried out by industry. One is concerned with a new weapon system for a helicopter, another is evaluating a new mine-hunting approach, and the third is still being worked up. All are due to report later this year. It is hoped that these will corroborate the findings so far.

5 Utilisation of PEAT

Three other uses for PEAT can be envisaged, utilising its capability to predict team performance (subject to further validation). These are:

- *Through-life capability management.* In order to meet service level agreements for delivered capability, it is already the case that maintenance teams will have to be created at relatively short notice to deal with some unexpected capability readiness problem. PEAT may be useful in ensuring a suitable team can be created.
- *Project cost estimation.* Tools such as COCOMO and COSYSMO make a number of not-always-explicit assumptions about the personnel involved in delivering a project. The inclusion of PEAT in these estimation techniques may help to improve these cost estimates by providing a more precise estimate of the performance of project personnel.
- *System health monitoring.* Current techniques concentrate heavily on the technology of the system, and tend to omit much consideration of the people actively present within the system. By including PEAT in these techniques it may be possible to provide a full, socio-technical assessment of the state of health of a system, not just a technological assessment.

6 Conclusions

In recognition of the fact that a further three independent tests of the tool are under way by industry

prior to its acceptance, conservative conclusions are in order

- A tool with some predictive power has been produced. However, it is not entirely clear how powerful the tool is, nor how extensive is its range of application
- It is believed to be the only tool able to provide estimates of team performance available for designers and human factors experts to use in the early stages of design.
- Planned developments to the user interface should make the tool usable by non-experts in human factors. This is believed to be an asset; if the engineer's own use of the tool shows that human factors issues must be addressed, this is likely to be a convincing argument.
- That the tool is in EXCEL spreadsheet form is good, but insufficient; it is necessary that a web-based version is available. This means that a more acceptable user interface could be given to the tool, and it can have a management process associated with it, both to maintain its integrity, to capture lessons from its use, and to develop the tool within the organisation hosting it. Some plans exist for this.

7 Acknowledgements

This tool has been developed within the ambience of an Engineering Doctorate. Thanks are therefore due to the Engineering and Physical Sciences Research Board and the sponsor, BAE Systems, without whose funding this would not have happened. Many, many others have provided insight and knowledge to this exercise; those most directly involved in this are Dr Siemieniuch, Prof Henshaw, and Prof. Haslam, all of Loughborough University, and Mr. Evans of BAE Systems. Their contributions are saluted.

8 References

1. Williams, J.C. *HEART – a proposed method for assessing and reducing human error.* in *9th Advances in Reliability Technology Symposium.* 1986. University of Bradford, UK.
2. Hollnagel, E., *Cognitive Reliability and Error Analysis Method (CREAM).* 1998, Den Haag: Elsevier Science.
3. Siegel, S., *Non-parametric statistics.* 1959, New York: McGraw-Hill.
4. Cochran, W.G., *Some methods for strengthening the common Chi-squared tests.* *Biometrics*, 1954. **10**: p. 417-451.